Asset Detection in Railroad Environments using Deep Learning-based Scanline Analysis

Johannes Wolf, Rico Richter and Jürgen Döllner
Hasso Plattner Institute, Faculty of Digital Engineering, University of Potsdam, Germany

Keywords: LiDAR, 3D Point Clouds, Digital Image Analysis, Semantic Classification.

Abstract: This work presents an approach for the automated detection of railroad assets in 3D point clouds from mobile mapping LiDAR scans using established convolutional neural networks for image analysis. It describes how images of individual scan lines can be generated from 3D point clouds. In these scan lines, objects such as tracks, signal posts, and axle counters can be detected using artificial neural networks for image analysis, previously trained on ground-truth data. The recognition results can then be transferred back to the 3D point cloud as a semantic classification result, or they are used to generate geometry or map data for further processing in GIS applications. Using this approach, trained objects can be found with high automation. Challenges such as varying point density, different data characteristics of scanning devices, and the massive amount of data can be overcome with this approach.

1 INTRODUCTION

Railway infrastructure is an essential backbone of today's transportation sector. Whether people or goods are being transported, trains serve as vehicles with exceptionally high transport capacity. To ensure safe operation, the infrastructure must be continuously monitored. Maintenance and repairs are time-consuming and costly, and breakdowns should be prevented as far as possible. For this purpose, it is advantageous to have up-to-date information on the railroad bed condition at all times (Chia et al., 2019).

While people often still have to walk along the tracks to check them for problems (Sanne, 2008), trains can be equipped with suitable hardware to collect digital information about the condition and update it continuously. For example, the tracks and ties' condition can be checked, and it can be determined whether the track or ballast is sagging at any point and whether the clearance area is being violated by growing vegetation. If this data is collected permanently, changes can be registered quickly, and the responsible authorities can react directly and take measures before a failure occurs (Ciocoiu et al., 2017).

Many national railroad companies, such as Deutsche Bahn in Germany and SBB in Switzerland, operate measuring trains that examine the tracks and their surroundings in detail during the journey (Wirth, 2008). Besides photo data, LiDAR scans are used for precise measurements, resulting in 3D point clouds.

3D point cloud analysis is done in different ways depending on multiple factors. Data can be acquired either with airborne systems or with mobile mapping vehicles at ground level. Furthermore, the data can be acquired either by LiDAR scans or via photogrammetry. This work focuses on processing LiDAR scans from mobile mapping scans in railroad environments.

Capturing with LiDAR is particularly suitable in the railroad environment because LiDAR is less dependent on environmental variables such as changing lighting in tunnels. On the other hand, photogrammetry is less suitable because, in many situations, it is not possible to generate a sufficient number of different perspectives on an object.

Figure 1 shows a train with LiDAR scanners mounted in front of the train. These scanners typically generate 3D point clouds with a scan line distance of 5 to 15 cm. The rotation of the laser beam during the measurement and simultaneous movement of the train result in a series of measurements in the form of a helix, in which each measuring point can be located by its distance from the scanner and the current angle of the laser. The points of one rotation of the laser are called a scan line. The resulting points can be visualized as a 3D point cloud of the entire track environment. Figure 2 shows a section of such a 3D point cloud. The individual scan lines, each of which are here 8 cm apart, are clearly visible.

Wolf, J., Richter, R. and Döllner, J.
Asset Detection in Railroad Environments using Deep Learning-based Scanline Analysis.
DOI: 10.5220/0010314704650470
In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2021) - Volume 4: VISAPP*, pages 465-470
ISBN: 978-989-758-488-6

Figure 1: Measurement train “Limez III” (Wirth, 2008).

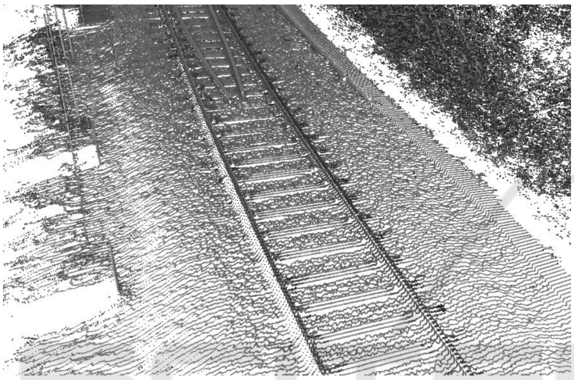


Figure 2: 3D point cloud from a railroad mobile mapping LiDAR scan, colored based on intensity values.

Automatic analysis of the data for early problem detection depends on the semantic classification of the information. Thus, points captured in the 3D point cloud must be grouped and identified in order to be able to derive statements about the nature of and distances between objects.

For this purpose, geometric analyses are often used, which analyze the properties of points and point groups to derive the most probable semantic class (Wolf et al., 2019b). In recent work, artificial neural networks such as PointNet++, which have learned to derive semantic classes from the structure of 3D point clouds, have also been delivering good results (Qi et al., 2017; Wang et al., 2019; Zhang et al., 2019).

The analysis of image data with Convolutional Neural Networks (CNNs), on the other hand, has been in use for a considerably longer time, and many techniques have been continuously improved over the years. Therefore, this paper presents a concept of how the scan line data of the LiDAR scans from railroad environments can be converted into image data to subsequently identify objects therein with established image analysis methods and use this information for further analyses.

2 RELATED WORK

Many analysis steps using 3D point clouds as input depend on using information about surfaces in the data since planar surfaces provide the basis for many other recognition steps. Approaches have been developed to perform a planar segmentation (Oehler et al., 2011). Plane recognition is even possible in sparse data, e. g., when lidar data from the environment of a moving car are acquired, as Wang et al. (2016) show.

Guan et al. (2016) compare how LiDAR information is used for road information inventory in various publications. For autonomous driving, the evaluation of LiDAR data in road space must be very fast so that the current environment can be evaluated immediately without delay. The use of CNNs for this purpose was investigated by Caltagirone et al. (2017). Specific objects in the road space, such as people, can also be detected in the data and used for safely controlling vehicles (Navarro-Serment et al., 2010). Many methods can also be transferred from the road to tracks. Stein et al. (2016) investigate how tracks of light rails can be detected automatically in LiDAR scans using variations in the distance values.

Arastounia (2017) points out the necessity of finding assets in the track environment and presents an algorithm for detecting rail tracks and contact cables by geometrical analysis of point positions based on an automated seed point search. Gézero and Antunes (2019) describe an approach to evaluate LiDAR point clouds of a rail environment using the angular information of a vertically mounted scanner at the front of the train. Along an imaginary line lying under the scanner, they determine the rails on either side of it and the ballast’s dimensions. In his doctoral thesis, Taheriandani (2016) describes approaches for detailed track analysis with LiDAR scanners that are directly aimed at the rails under the train to detect the smallest deviations. Shang et al. (2018) present an approach for finding rail defects by using CNNs on railway image data.

Detecting structures and objects in images is a relevant research field for many applications, such as face recognition, license plate identification, or medical imagery analysis. U-Net, initially developed for the medical sector, is now widely used in image segmentation (Ronneberger et al., 2015). With this network’s help, specific areas in images can be recognized with pixel accuracy, such as cancer cells and streets in aerial images (Zhang et al., 2018). Another promising implementation is YOLO (You only look once), which only returns labeled bounding boxes, but can process the provided images very fast (Redmon and Farhadi, 2018). In a railway context, Yanan et al.

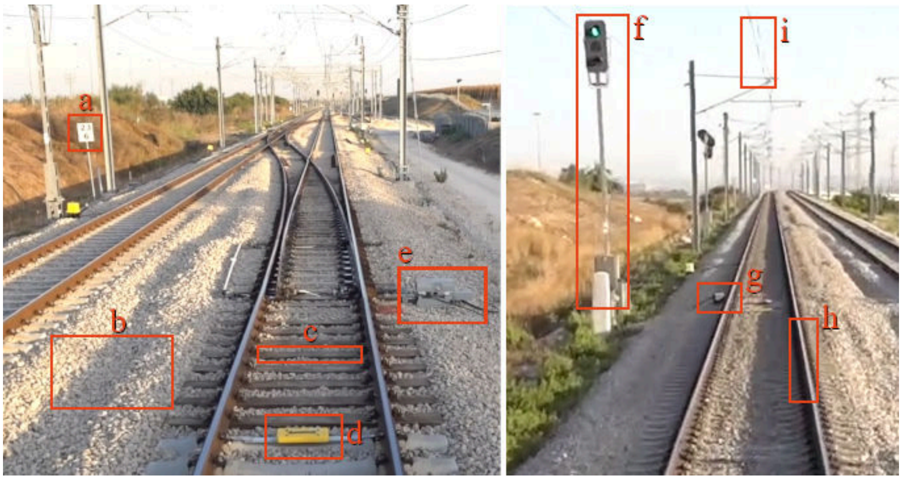


Figure 3: Typical objects found in railroad environments: Sign (a), ballast (b), tie (c), balise (d), switch motor (e), signal pole (f), axle counter (g), rail (h), catenary (i).

(2018) use YOLO to detect problems on the surface of rails, Yang et al. (2020) use it to recognize pole numbers in images.

Hu and Ye (2013) use scan lines in aerial data to detect buildings while Yan et al. (2016) present an approach using them to recognize road markings in mobile mapping data.

3 CONCEPT

The idea of the approach presented in this paper is based on the assumption that the performance of 2D image recognition can be utilized due to the nature of the data when individual scan lines are available. Thus, even three-dimensional objects can be precisely and efficiently identified and semantically classified via this approach. Typical objects that should be classified in railroad data are shown in Figure 3.

When rendering images from 3D point cloud data for object classification, positioning the virtual camera is of utmost importance. As described in previous work (Wolf et al., 2019a), top-down views are suitable for detecting objects such as road markings and utility hole covers in mobile mapping data. However, having catenaries and signal bridges above railroad tracks, these are occluding essential parts of the track in a top-down view and are therefore hindering comprehensive classification. Furthermore, in tunnels, catenaries and signals are often mounted to the ceiling, making it very difficult to do a top-down view analysis. Using the scan lines is an obvious choice because all objects are captured from the train's perspective, and therefore everything necessary is visible in the data.

First, individual scan lines are identified: At best, the measurement data already contain information about the point's position within the individual scan lines by their sequence or timestamp. In this case, all points can be combined within a run of possible scan angles from 0 to 360 degrees, with 0 degrees being straight above the train, and as soon as the angle of the following point jumps back, a new scan line begins. In a prototypic implementation, about 21 million points per second were segmented this way into individual scan lines from an ordered 3D point cloud.

If this information is not available in the current data set, the scan lines can also be derived in an additional preparatory step: Along the measurement data's trajectory, scan lines are generated perpendicular to it by grouping adjacent points into scan lines. Here, approximate solutions are sufficient because the affiliation to a particular scan line is not decisive for the later analysis.

Now all scan lines can be rendered individually as 2D images. A particular color (e. g., white (255)) is used as masking for the areas not containing data, and all other pixels can be colored with gray levels according to the intensity of the measuring points at the respective position, mapping the lowest intensity to black (0), the highest to almost white (254). An additional image channel containing the rendered points' IDs will also be included so that the result can be mapped back into the point cloud after the image classification. If several points are rendered in the same pixel, the last rendered point's ID is stored.

Figure 4 shows the difference of the surface's smoothness between two rails depending on whether a tie is placed at this position or the ballast is exposed. Thus, ties can be identified if the scan lines are placed at suitable distances from each other.

(a) Two rails with a tie in between.

(b) Two rails with ballast in between.

Figure 4: Two scan lines of a railroad track on a ballast hill. Grayscale represents the intensity values of the points.

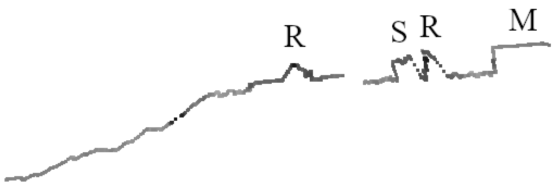


Figure 5: Scan lines of two rails (R), a switch tongue (S) and a box containing the motor (M) next to the rails.

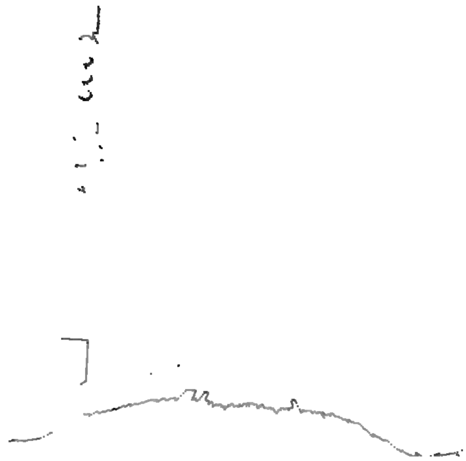
A large variety of objects can be found in the immediate neighborhood of the rails. Figure 5 shows a scan line at the beginning of a track switch, where the switch tongue is placed close to one of the rails and a box with the switch motor on the other side of the rail. The box shadows the area on the right side, so this part of the ballast hill is missing.

Figure 6 shows two scan lines of a signal post next to the rails. Due to its shape, the front of the actual signal can be seen in multiple scan lines before the signal post. In total, the structure has a length of approximately 1.5 meters.

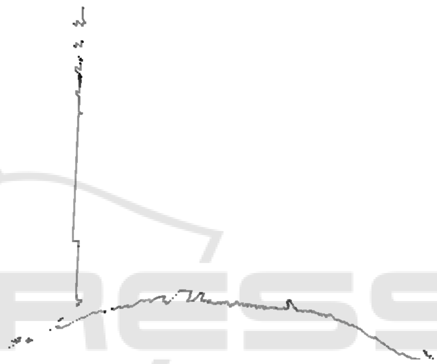
Possible semantic classes analyzed with this technique are rails (running rails, guard rails, switch tongues), ties, ballast, catenary posts, signals, signal posts, and signal gantries.

A large number of rendered images will be needed for training purposes. This data could be generated manually by labeling individual pixels and bounding boxes within the scan line images. A faster approach would be using pre-classified 3D point clouds (which have been created either manually or by a different automated approach) so that the semantic information can already be included when rendering the 2D images.

Suitable networks for the analysis of the rendered images are, for example, U-Net and YOLO. Both follow a different approach but could provide similarly relevant results for the application described here. While U-Net classifies individual pixels, YOLO only determines bounding boxes for recognized objects. However, since there are hardly any overlaps of



(a) Front part of the signal.



(b) Signal post, six scan lines behind the one above.

Figure 6: Scan lines of a signal post next to two rails. An axle counter is attached to the outside of the left rail.

objects in individual scan lines and the objects to be found, such as rails, ties, and signals, can be covered relatively well by rectangles, this result should also be sufficient. When using YOLO, all non-background pixels within the bounding box of a recognized object could get assigned the corresponding semantic class, and then they would be treated similarly to the images classified pixel by pixel with U-Net.

Figures 7 and 8 show exemplary results of the semantic classification with YOLO and U-Net on a scan line.

Once the semantic class for each pixel is determined, the information can be mapped back into the 3D point cloud by using the ID channel. In case the point density is higher than the resolution of the rendered images, several points have been covered by the same pixel. In this case, all points in the immediate neighborhood of the point just classified can also receive the respective semantic class so that all points will receive semantic information.

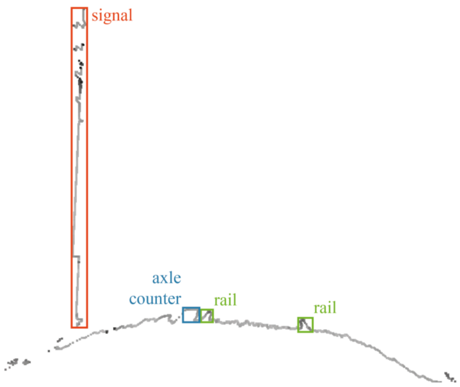


Figure 7: Exemplary semantic classification results of the scan line shown in Figure 6b using YOLO. Bounding boxes are placed around identified objects.

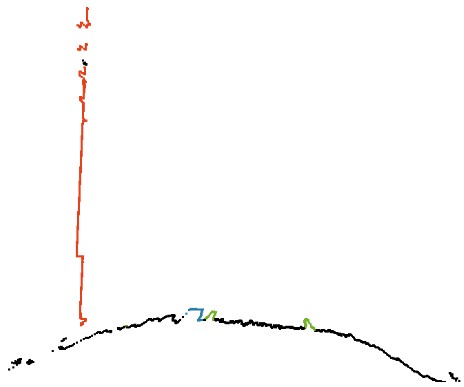


Figure 8: Exemplary semantic classification results of the scan line shown in Figure 6b using U-Net. Points are colored based on semantic class: Signal (red), rail (green), axle counter (blue), other (black).

Processing the data results in a 3D point cloud with semantic information attached to each point. This data can then be used for the previously described use cases.

Simple post-processing steps and plausibility checks can further improve the results of the classification. For example, axle counters must lie close to the track, and rails of a track always run parallel with a fixed, previously known distance between them. Such conditions can be checked for after the classification. For example, objects identified as axle counters but not located right next to a track could then be discarded and, e. g., classified as "other".

4 CONCLUSION AND FUTURE WORK

First prototypical tests show that the described approach is suitable for the semantic classification of 3D point clouds of railroad environments. Individual scan lines can be analyzed by rendering images and using established image analysis for the classification. However, the performance is still to be determined in more extensive tests. It should also be evaluated whether using one-dimensional CNNs as they are used, e. g., for sound classification or movement recognition (Cho and Yoon, 2018; Abdoli et al., 2019) can perform more efficient on the given task.

The 3D point clouds classified by this approach can then be used for various tasks in track maintenance. For example, location maps could be generated for the detected objects in the track area, or existing data could be compared with the information obtained here and adjusted if needed.

The approach can be extended in several ways to presumably further improve the results. The generated images could be centered along the train's trajectory, so the position within the images provides information about the objects displayed. For example, the rails would then always be found in a similar position. Furthermore, several scan lines could be viewed simultaneously to enrich the images with context information. For this purpose, the images could be given additional layers so that, for example, three or five scan lines are contained in one image, and the preceding and following scan lines have additional influence on the scan line to be analyzed.

REFERENCES

Abdoli, S., Cardinal, P., and Koerich, A. L. (2019). End-to-end environmental sound classification using a 1d convolutional neural network. *Expert Systems with Applications*, 136:252–263.

Arastounia, M. (2017). An enhanced algorithm for concurrent recognition of rail tracks and power cables from terrestrial and airborne lidar point clouds. *Infrastructures*, 2(2):8.

Caltagirone, L., Scheidegger, S., Svensson, L., and Wahde, M. (2017). Fast lidar-based road detection using fully convolutional neural networks. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 1019–1024. IEEE.

Chia, L., Bhardwaj, B., Lu, P., and Bridgelall, R. (2019). Railroad track condition monitoring using inertial sensors and digital signal processing: A review. *IEEE Sensors Journal*, 19:25–33.

Cho, H. and Yoon, S. M. (2018). Divide and conquer-based 1d cnn human activity recognition using test data sharpening. *Sensors*, 18(4):1055.

Ciocoiu, L., Siemieniuch, C. E., and Hubbard, E.-M. (2017). From preventative to predictive maintenance: The organisational challenge. *Proceedings of the In-*

stitution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 231(10):1174–1185.

Gézero, L. and Antunes, C. (2019). Automated three-dimensional linear elements extraction from mobile lidar point clouds in railway environments. *Infrastructures*, 4(3):46.

Guan, H., Li, J., Cao, S., and Yu, Y. (2016). Use of mobile lidar in road information inventory: A review. *International Journal of Image and Data Fusion*, 7(3):219–242.

Hu, X. and Ye, L. (2013). A fast and simple method of building detection from lidar data based on scan line analysis. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 7–13.

Navarro-Serment, L. E., Mertz, C., and Hebert, M. (2010). Pedestrian detection and tracking using three-dimensional ladar data. *The International Journal of Robotics Research*, 29(12):1516–1528.

Oehler, B., Stueckler, J., Welle, J., Schulz, D., and Behnke, S. (2011). Efficient multi-resolution plane segmentation of 3d point clouds. In *International Conference on Intelligent Robotics and Applications*, pages 145–156. Springer.

Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems*, pages 5099–5108.

Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Sanne, J. M. (2008). Framing risks in a safety-critical and hazardous job: risk-taking as responsibility in railway maintenance. *Journal of Risk Research*, 11:645 – 658.

Shang, L., Yang, Q., Wang, J., Li, S., and Lei, W. (2018). Detection of rail surface defects based on cnn image recognition and classification. *2018 20th International Conference on Advanced Communication Technology (ICACT)*, pages 45–51.

Stein, D., Spindler, M., Kuper, J., and Lauer, M. (2016). Rail detection using lidar sensors. *International Journal of Sustainable Development and Planning*, 11(1):65–78.

Taheriandani, M. (2016). *The Application of Doppler LIDAR Technology for Rail Inspection and Track Geometry Assessment*. PhD thesis, Virginia Tech.

Wang, L., Meng, W., Xi, R., Zhang, Y., Ma, C., Lu, L., and Zhang, X. (2019). 3d point cloud analysis and classification in large-scale scene based on deep learning. *IEEE Access*, 7:55649–55658.

Wang, W., Sakurada, K., and Kawaguchi, N. (2016). Incremental and enhanced scanline-based segmentation method for surface reconstruction of sparse lidar data. *Remote Sensing*, 8(11):967.

Wirth, H. (2008). Der neue lichtraummesszug limez iii der deutschen bahn ag. *Zeitschrift für Geodäsie, Geoinformation und Landmanagement*, 133(3):180–186.

Wolf, J., Richter, R., Discher, S., and Döllner, J. (2019a). Applicability of neural networks for image classification on object detection in mobile mapping 3d point clouds. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(4/W15):111–115.

Wolf, J., Richter, R., and Döllner, J. (2019b). Techniques for automated classification and segregation of mobile mapping 3d point clouds. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 201–208.

Yan, L., Liu, H., Tan, J., Li, Z., Xie, H., and Chen, C. (2016). Scan line based road marking extraction from mobile lidar point clouds. *Sensors*, 16(6):903.

Yanan, S., Hui, Z., Li, L., and Hang, Z. (2018). Rail surface defect detection method based on yolov3 deep learning networks. In *2018 Chinese Automation Congress (CAC)*, pages 1563–1568. IEEE.

Yang, Y., Zhang, W., He, Z., and Li, D. (2020). High-speed rail pole number recognition through deep representation and temporal redundancy. *Neurocomputing*, 415:201–214.

Zhang, Z., Hua, B.-S., and Yeung, S.-K. (2019). Shellnet: Efficient point cloud convolutional neural networks using concentric shells statistics. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1607–1616.

Zhang, Z., Liu, Q., and Wang, Y. (2018). Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753.